

Variaciones genómicas y expresión génica

Valencia, Septiembre 2011
CEFIRE

Francisco García

fgarcia@cipf.es

<http://bioinfo.cipf.es/fgarcia>

Departamento de Bioinformática y Genómica
Centro de Investigación Príncipe Felipe (CIPF)
Valencia (España)



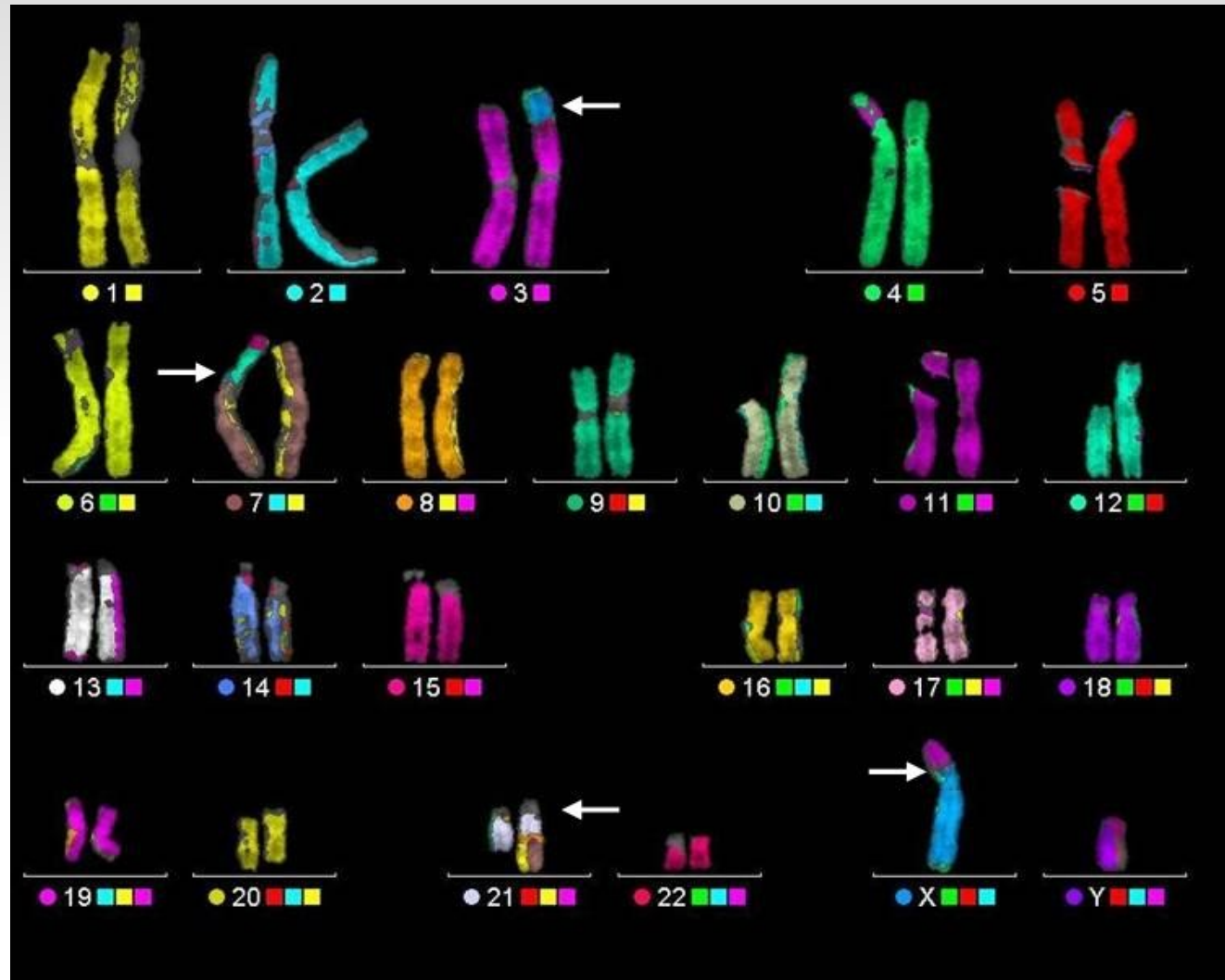
Información biológica

ADN → **ARN** → **proteína**

Genoma → **Transcriptoma** → **Proteoma**

Variaciones genómicas

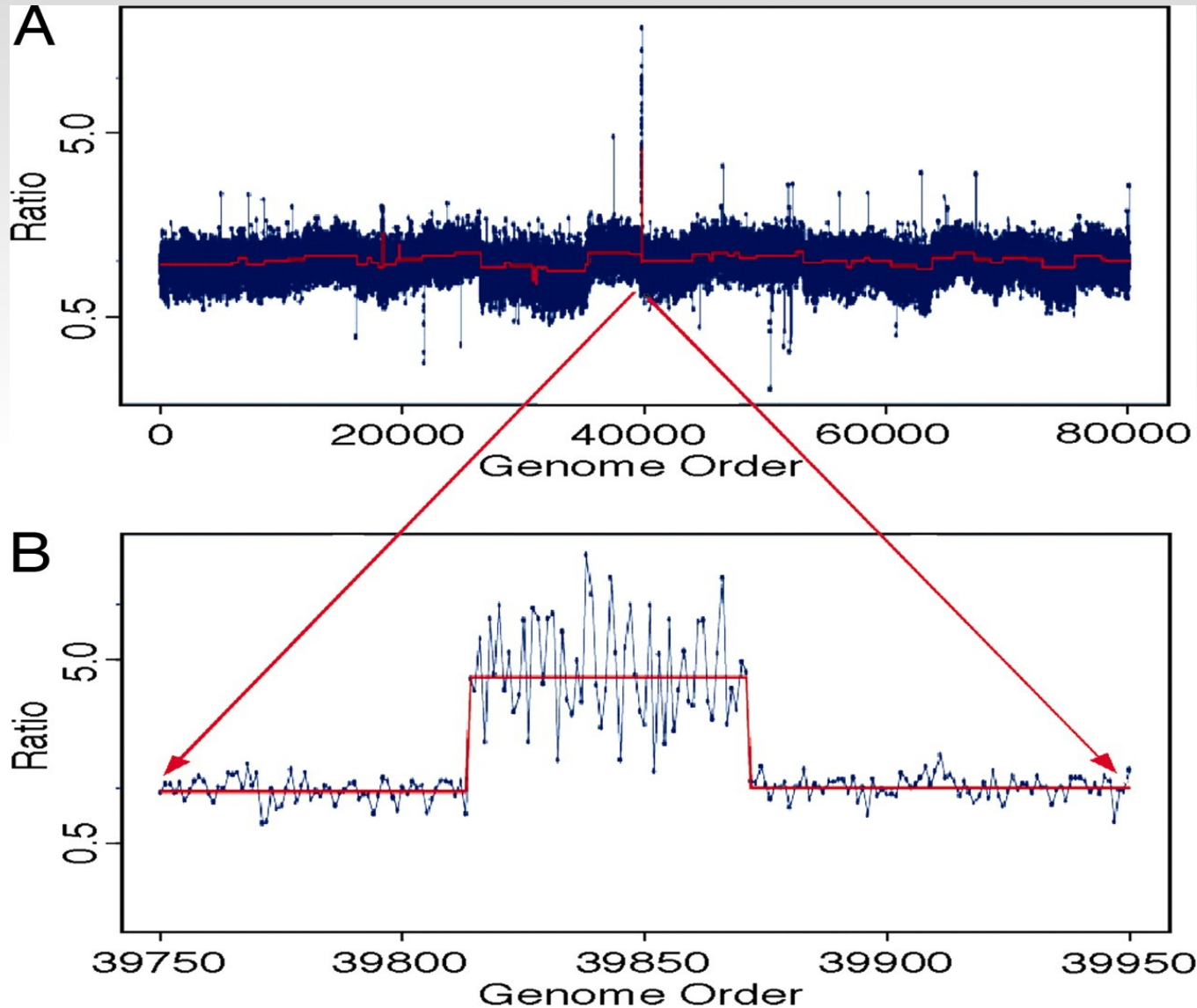
Cariotipo



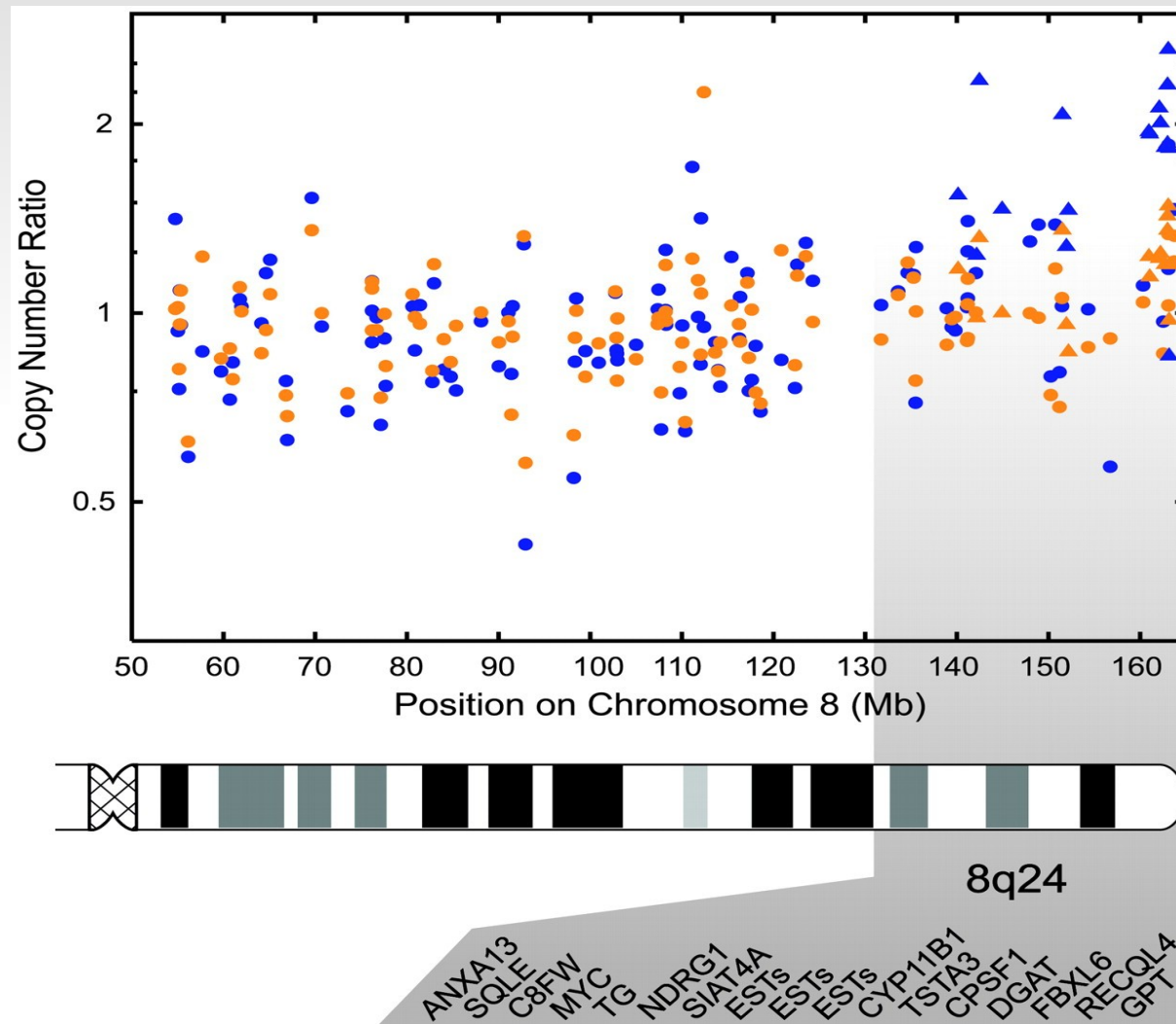
Tipos de variaciones genómicas

- Variación del número de copias (CNV)
 - Duplicidad – Amplificación
 - Delección
- Polimorfismos
 - Polimorfismo de un solo nucleótido (SNPs)

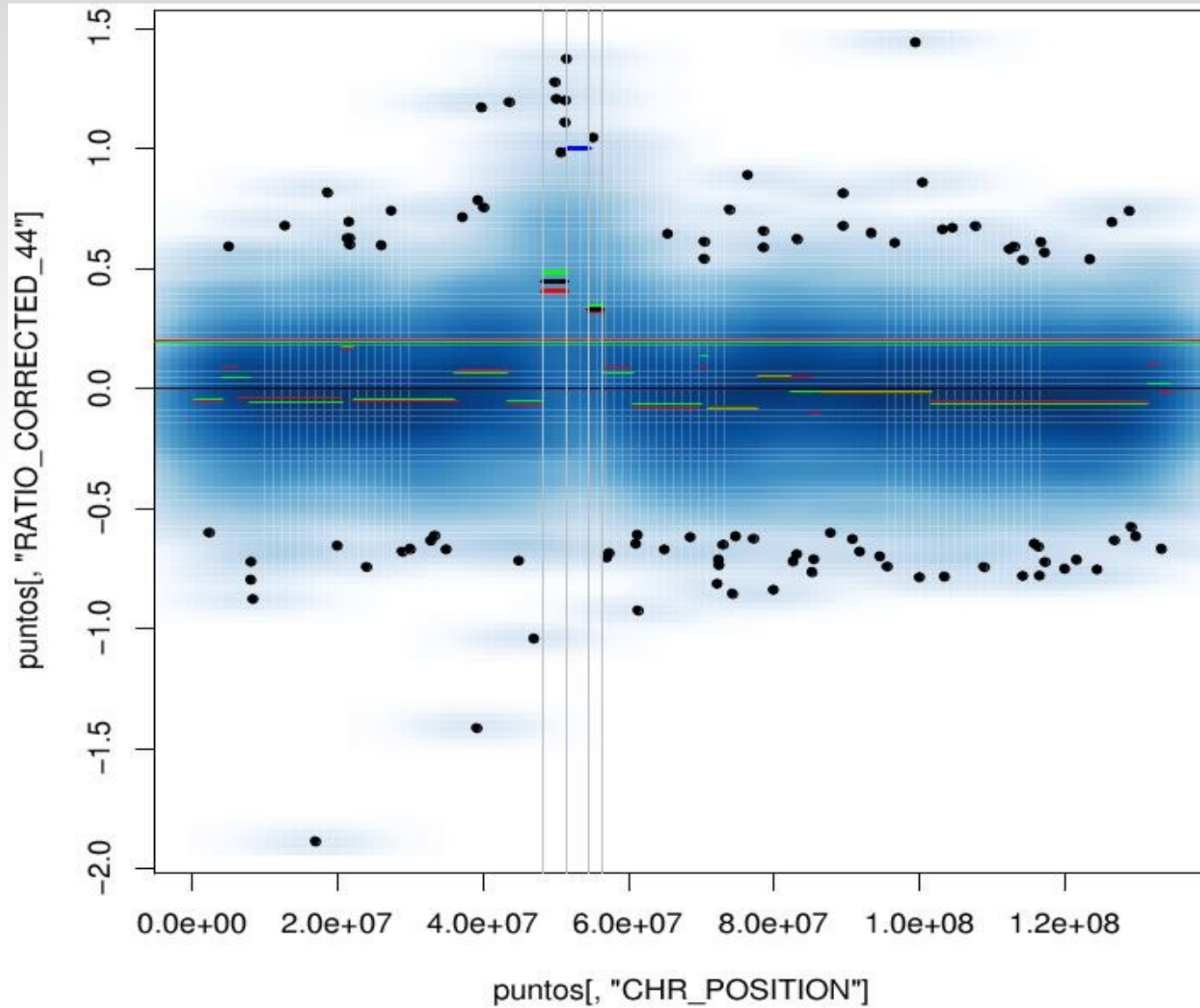
Análisis de segmentación



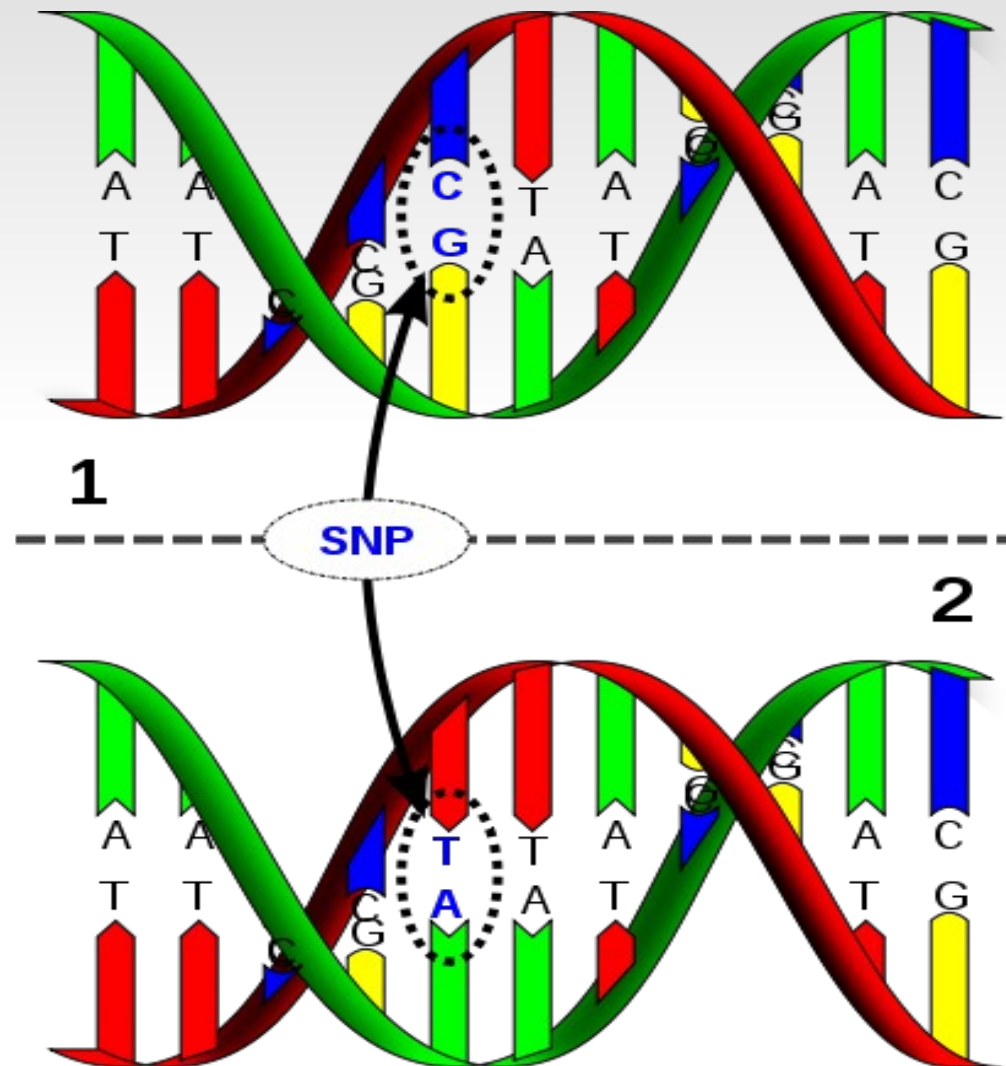
Combinando datos de varias muestras



Búsqueda de regiones comunes



Single nucleotide polymorphisms (SNP)



Estudios de asociación de SNPs

individuo **SNP1** **SNP2** Enfermedad

array1	A/A	A/B	caso
array2	A/B	A/A	caso
array3	A/B	A/B	caso
array4	A/A	A/B	caso
array5	A/B	A/B	control
array6	A/B	A/A	control
array7	B/B	A/B	control
array8	A/A	A/B	control
array9	B/B	A/B	control
array10	B/B	A/A	control
array11	B/B	A/B	control
array12	B/B	A/B	control

	A	B
caso	6	2
control	4	8

$$\frac{6 \cdot 8}{2 \cdot 4} > 1$$

→ Alelo A
asociado a caso

Estratificación

Población 1

	A	B
caso	2	1
control	2	1

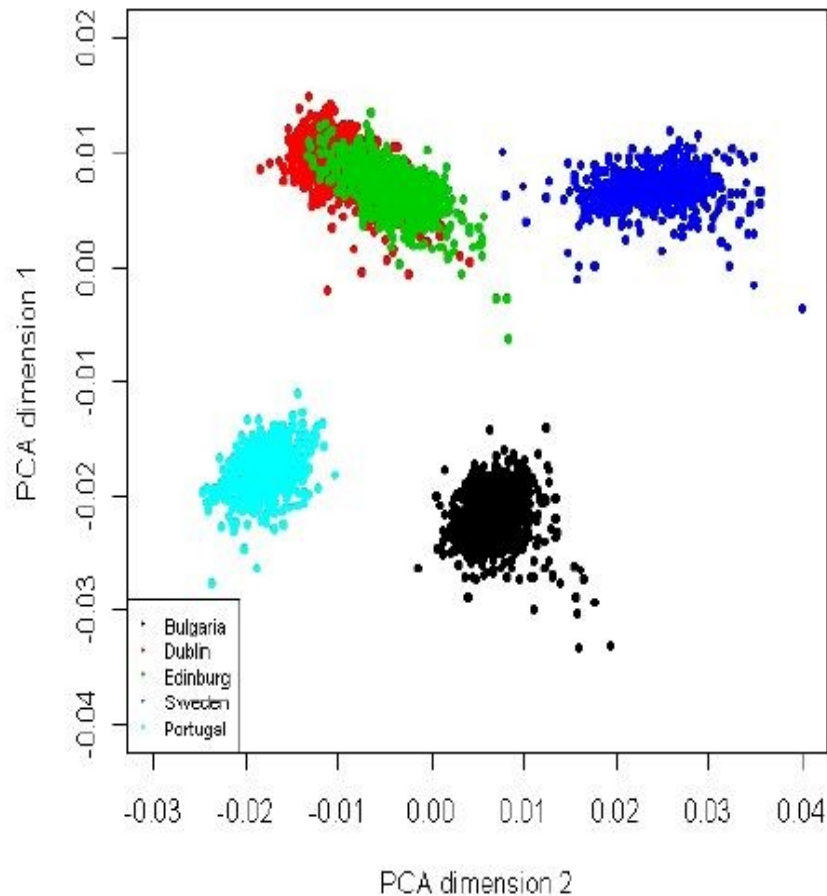
$$\longrightarrow \frac{2 \cdot 1}{2 \cdot 1} = 1$$

Población 2

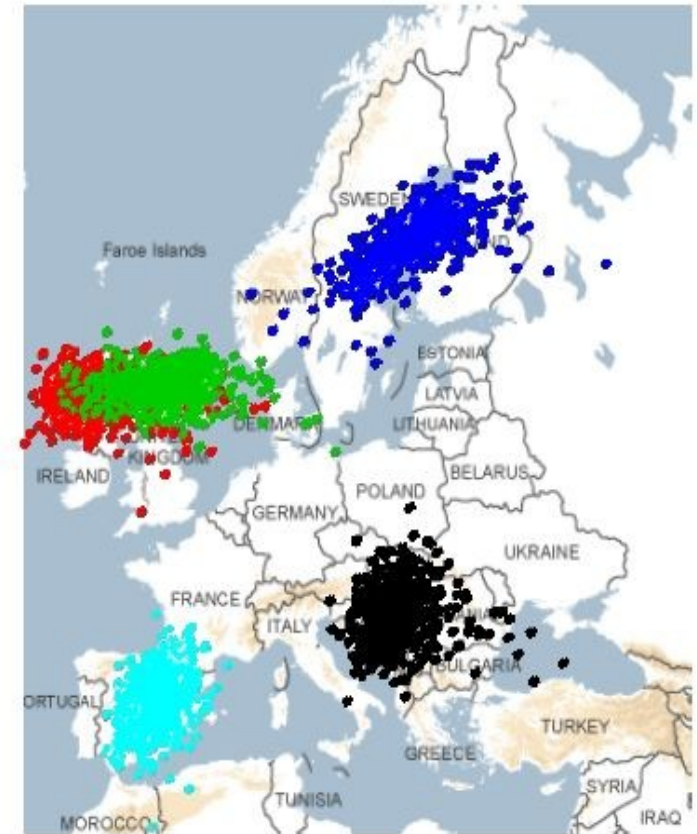
	A	B
caso	4	1
control	2	7

$$\frac{4 \cdot 7}{2 \cdot 1} > > 1$$

Estratificación de la población



a.

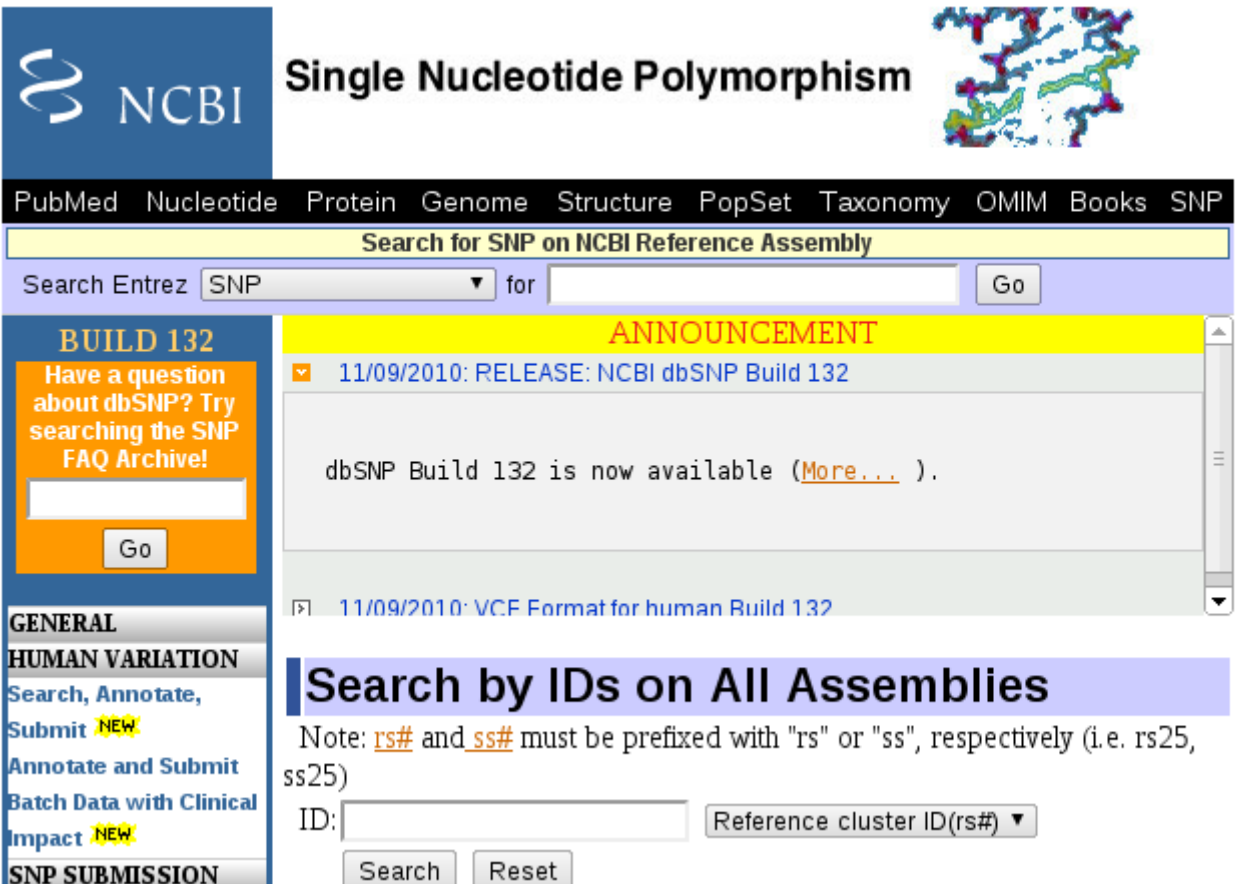


b.

Valentina Moskvina et al.
Human Heredity Vol. 70, No. 2, 2010

Recursos - Bases de datos

- Búsqueda, anotación e inclusión de SNPs



The screenshot displays the NCBI SNP database homepage. At the top, the NCBI logo is on the left, and the text 'Single Nucleotide Polymorphism' is in the center, with a colorful molecular structure graphic on the right. Below this is a navigation bar with links: PubMed, Nucleotide, Protein, Genome, Structure, PopSet, Taxonomy, OMIM, Books, and SNP. A search bar is present with the text 'Search for SNP on NCBI Reference Assembly'. Below the search bar, there is a section for 'BUILD 132' with a message: 'Have a question about dbSNP? Try searching the SNP FAQ Archive!' and a 'Go' button. To the right of this is an 'ANNOUNCEMENT' section with a yellow header. It contains two items: a checked box next to '11/09/2010: RELEASE: NCBI dbSNP Build 132' and a text block stating 'dbSNP Build 132 is now available (More...)'. Below this is another announcement: '11/09/2010: VCF Format for human Build 132'. On the left side, there is a sidebar with a 'GENERAL' section containing links for 'HUMAN VARIATION', 'Search, Annotate, Submit NEW', 'Annotate and Submit', 'Batch Data with Clinical Impact NEW', and 'SNP SUBMISSION'. At the bottom, there is a section titled 'Search by IDs on All Assemblies' with a note: 'Note: rs# and ss# must be prefixed with "rs" or "ss", respectively (i.e. rs25, ss25)'. Below the note is a search form with an 'ID:' label, a text input field, a 'Reference cluster ID(rs#)' dropdown menu, and 'Search' and 'Reset' buttons.

NCBI Single Nucleotide Polymorphism

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Books SNP

Search for SNP on NCBI Reference Assembly

Search Entrez SNP for Go

BUILD 132

Have a question about dbSNP? Try searching the SNP FAQ Archive!

Go

ANNOUNCEMENT

11/09/2010: RELEASE: NCBI dbSNP Build 132

dbSNP Build 132 is now available ([More...](#)).

11/09/2010: VCF Format for human Build 132

GENERAL

HUMAN VARIATION

Search, Annotate, Submit **NEW**

Annotate and Submit

Batch Data with Clinical Impact **NEW**

SNP SUBMISSION

Search by IDs on All Assemblies

Note: **rs#** and **ss#** must be prefixed with "rs" or "ss", respectively (i.e. rs25, ss25)

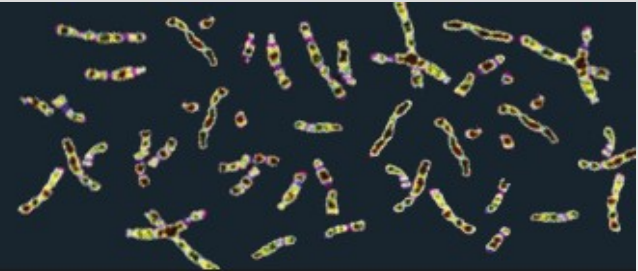
ID: Reference cluster ID(rs#)

Search Reset

Recursos- Variaciones individuales

1000 Genomes

A Deep Catalog of Human Genetic Variation



[Home](#) [About](#) [Data](#) [Analysis](#) [Participants](#) [Contact](#) [Browser](#) [Wiki](#) [FTP search](#)

LATEST ANNOUNCEMENTS

THURSDAY JUNE 23, 2011

June 2011 Data Release

Genotypes for 1094 individuals for the [May 2011 snp calls](#) from the 20101123 sequence and alignment release of the 1000 genomes project has now been made. This release is based on the GRCh37 assembly of the human genome and are released in the format [VCF 4.0](#)

Our [FAQ](#) contains instructions on how to get [smaller subsections](#) of these files

Data access links: [EBI / NCBI](#)

Link to additional information: [README file](#)

THURSDAY MAY 12, 2011

May 2011 Data Release

Full Project low coverage SNP call release

SNP calls based on 1094 individuals from the 20101123 sequence and alignment release of the 1000 genomes project has now been made. This release is based on the GRCh37 assembly of the human genome and are released in the format [VCF 4.0](#)

NAVIGATION

- [Frequently Asked Questions](#)

LINKS



[All Project Announcements](#)



[Sample and Project Information](#)




[Media Archive](#)



[Download the 1000 Genomes Pilot Paper](#)

Recursos- Mapas de haplotipos



International HapMap Project

[Home](#) | [About the Project](#) | [Data](#) | [Publications](#) | [Tutorial](#)

[中文](#) | [English](#) | [Français](#) | [日本語](#) | [Yoruba](#)

The International HapMap Project is a partnership of scientists and funding agencies from Canada, China, Japan, Nigeria, the United Kingdom and the United States to develop a public resource that will be available to the scientific community and pharmaceuticals. See "[About the International HapMap Project](#)" for more information.

Project Information

- [About the Project](#)
- [HapMap Publications](#)
- [HapMap Tutorial](#)
- [HapMap Mailing List](#)
- [HapMap Project Participants](#)

Project Data

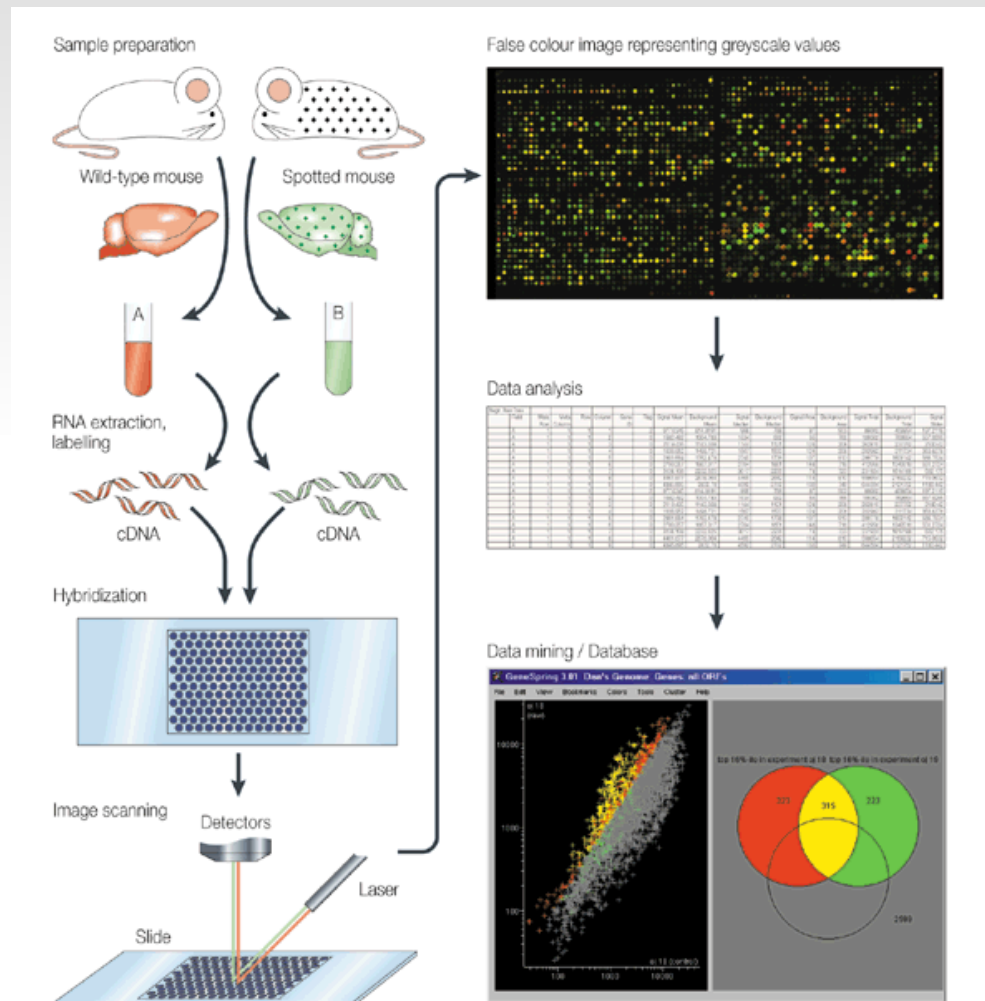
- [HapMap Genome Browser release #28 \(Phases 1, 2 & 3 - merged genotypes & frequencies \)](#)
- [HapMap3 Genome Browser release #3 \(Phase 3 - genotypes & frequencies \)](#)
- [HapMap Genome Browser release #27 \(Phase 1, 2 & 3 - merged genotypes & frequencies \)](#)
- [HapMap3 Genome Browser release #2 \(Phase 3 - genotypes, frequencies & LD \)](#)
- [HapMap Genome Browser release#24 \(Phase 1 & 2 - full dataset \)](#)
- [GWAs Karyogram](#)
- [HapMart](#)

News

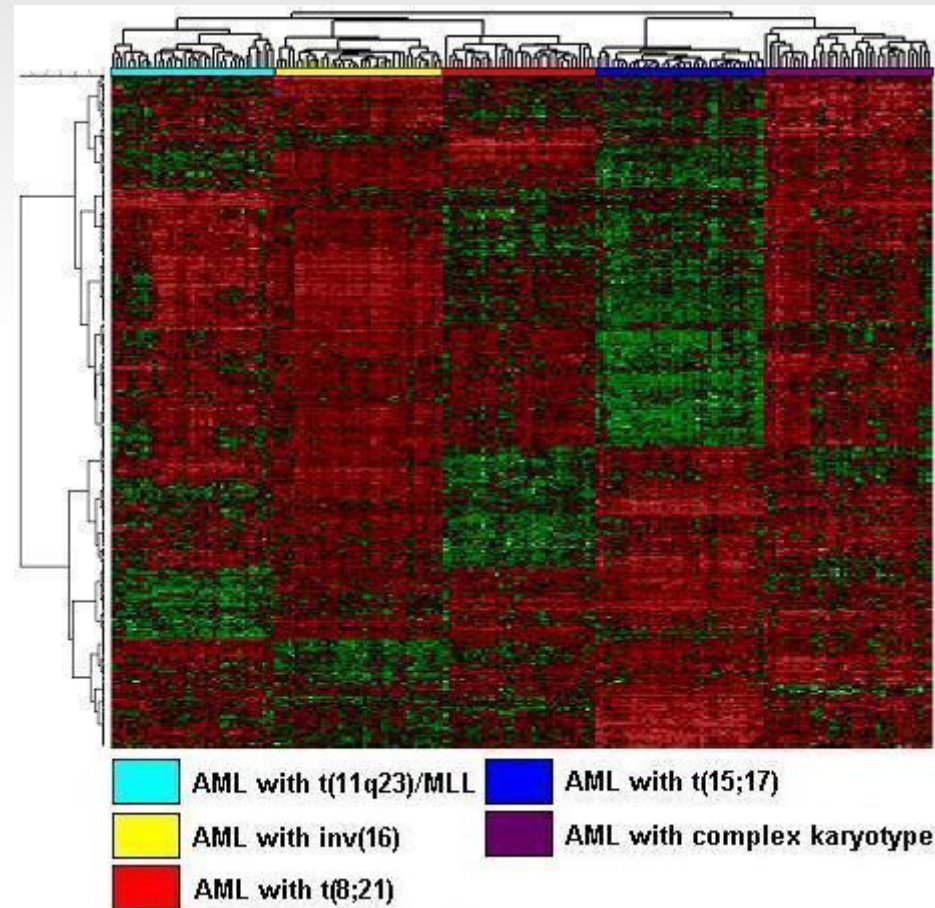
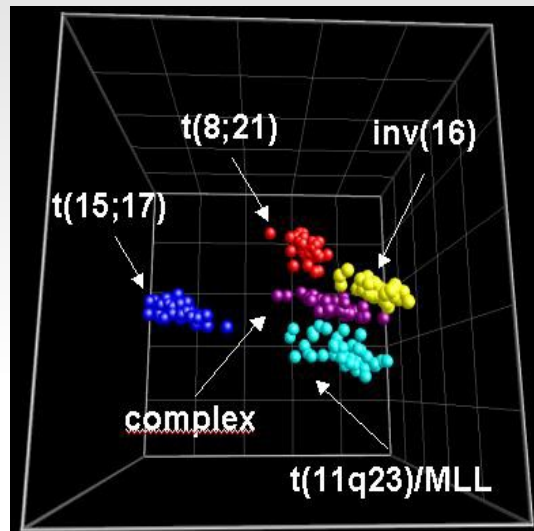
- 2011-06-13: **HapMap help desk announcement**
There was a problem with the HapMap help desk system. In the past several weeks, emails sent to hapmap-help@ncbi.nlm.nih.gov were not being received. Please resend your email request if you sent emails to the HapMap help desk in the past several weeks. Sorry for the inconvenience.
- 2011-04-20: **Hapmap help desk service interruption notice**
There will be no help desk support from 05/03/2011 to 05/23/2011. Sorry for the inconvenience.
- 2011-02-02: **Haploview issues with rel 28 data**
Recently, there are several questions about Haploview data format errors when users tried to analyze HapMap release 28. Haploview is developed and maintained by an organization different from HapMap. Please contact Haploview help desk (haploview@biodigital.com) for more information.
- 2011-01-19: **HapMap phase II recombination rate on GRCh37**
The liftover of the HapMap II genetic map from human genome build b35 to GRCh37 is available. Data is [available for download](#).
- 2010-08-18: **HapMap Public Release #28**
Genotypes and frequency data in hapmap format are now available for data in merged HapMap phases I+II+III release #28.

Estudios de expresión génica

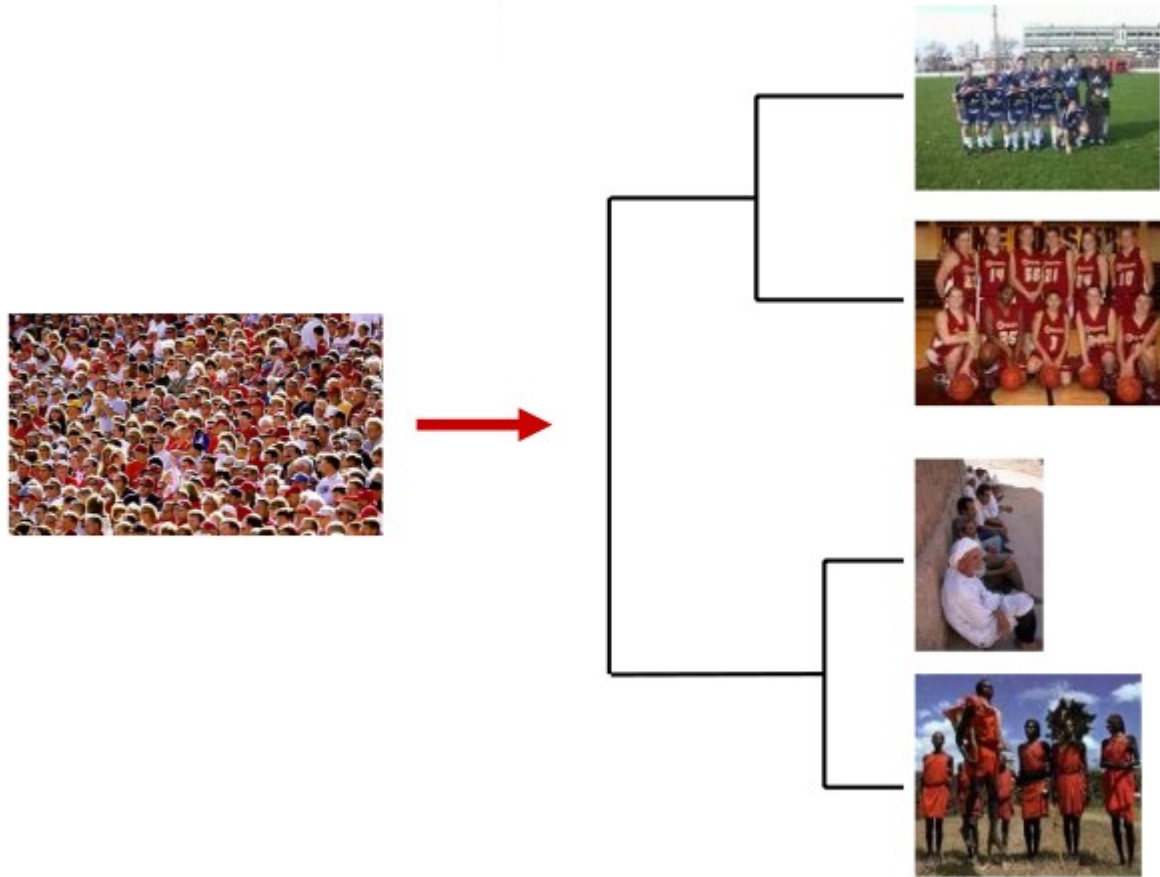
Comparación de clases



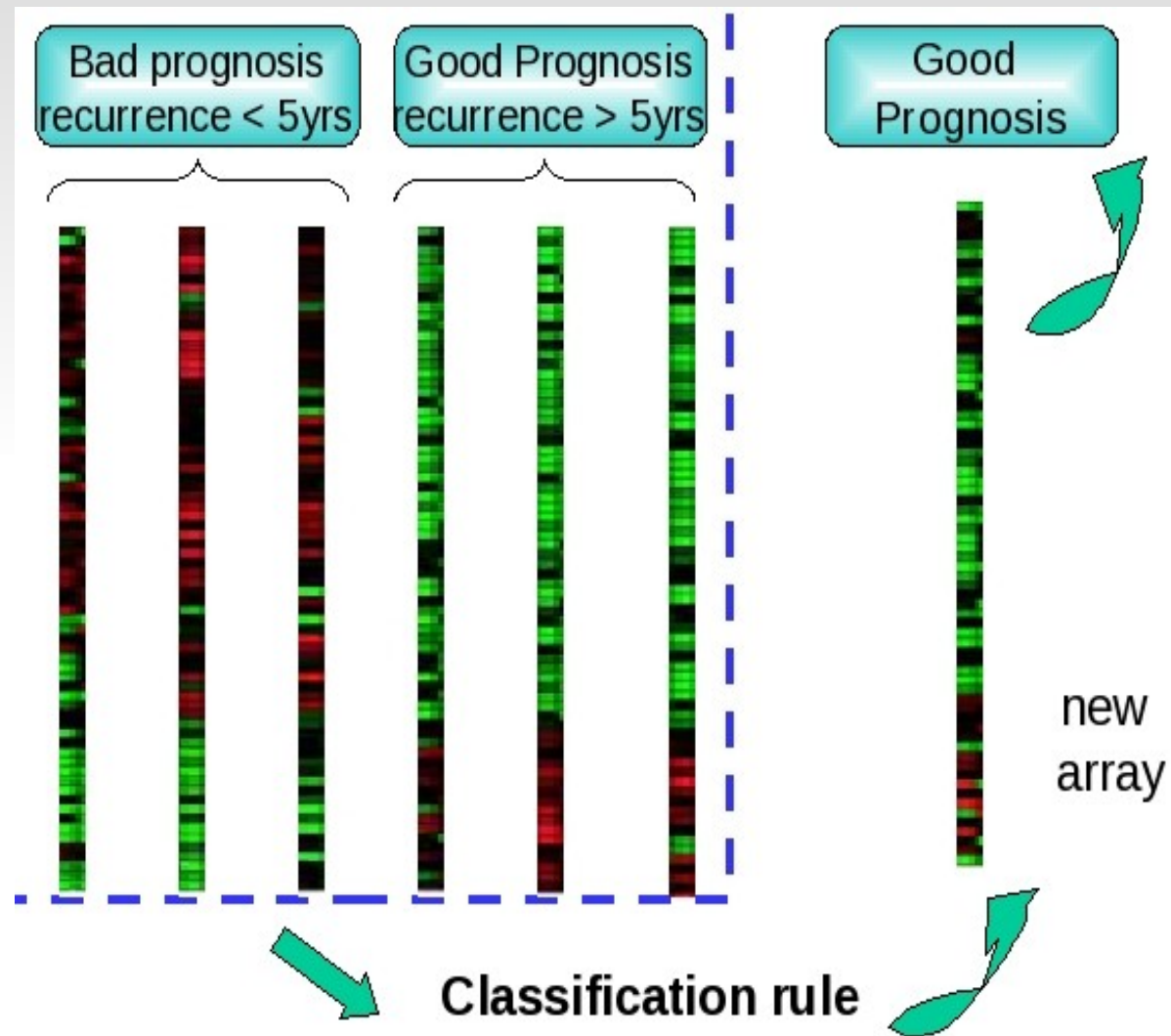
Descubrimiento de clases



Clustering



Predicción de clases



Análisis de datos de expresión



Análisis de datos de expresión

Preprocesamiento

- Expresión diferencial
- Clustering
- Predictores

Análisis
Funcional

**Regulatory
elements**

miRNA, CisRed
Transcription Factor
Binding Sites

Gene Ontology

Biological Process
Molecular Function
Cellular Component

**KEGG
pathways**

**Bases de datos
biológicos**

Análisis de datos de expresión

Carga de
datos

Preprocesamiento

- Expresión
diferencial
- Clustering
- Predictores

Análisis
Funcional

**Regulatory
elements**

miRNA, CisRed
Transcription Factor
Binding Sites

Gene Ontology

Biological Process
Molecular Function
Cellular Component

**KEGG
pathways**

**Bases de datos
biológicos**

Babelomics

home | help | tutorial | contact

BABELOMICS4

gene expression and functional profiling analysis suite

Upload data

Processing

Expression

Genomic

Functional analysis

Utilities

anonymous working on project default 0 Kb of 1.00 Gb (0.00%) no active jobs

logout

Expression

- Differential expression
 - Class comparison:

Study differential expression among more two or more array classes....read more
 - Correlation

Study expression among more two or more array classes....read more
 - Survival

Study the relationships between the expression of the genes and the survival time of the cells.
 - Time/dosage series

The module finds genes with a changing pattern along time or increasing dose concentrations, ...read more
- Predictors
 - Class prediction

Builds prediction rules and allows using them for further sample classification.
 - Clustering

These methods use (implicitly or explicitly) a distance function and an algorithm to join together genetic elements...read more
 - Biclustering

Integrated environment for biclustering analysis of time series gene expression data

Job list

no jobs found

Ejemplo de análisis con Babelomics

Descripción de los datos

- 24 muestras correspondientes a 12 individuos entrenados y 12 sedentarios
- Arrays Affymetrix, 22.000 genes

¿Qué queremos hacer?



- Análisis de expresión diferencial

Ejemplo de análisis con Babelomics

Formato de los datos

#VARIABLE	entrenado	entrenado	sedentario	sedentario	sedentario	entrenado
#NOMBRES	ARRAY1	ARRAY2	ARRAY3	ARRAY4	ARRAY5	ARRAY6
AFFX-BioB-5_at	7.521	6.933	6.659	7.589	6.347	7.623
AFFX-BioB-M_at	8.394	7.695	7.722	8.404	7.055	8.487
AFFX-BioB-3_at	8.163	7.610	7.545	8.113	7.207	8.176
AFFX-BioC-5_at	8.959	8.367	8.269	9.160	7.940	9.190
AFFX-BioC-3_at	8.381	7.679	7.713	8.551	7.568	8.682
AFFX-BioDn-5_at	9.173	8.245	8.314	9.009	8.766	8.975
AFFX-BioDn-3_at	11.918	11.398	11.353	11.726	11.512	11.797
AFFX-CreX-5_at	12.275	11.640	11.781	12.007	11.685	12.003
AFFX-CreX-3_at	12.919	12.274	12.306	12.482	12.343	12.567
AFFX-DapX-5_at	3.649	3.625	3.639	3.500	3.746	3.714
AFFX-DapX-M_at	3.879	3.907	3.864	3.926	3.879	3.854
AFFX-DapX-3_at	3.558	3.602	3.550	3.562	3.627	3.599
AFFX-LysX-5_at	3.546	3.662	3.614	3.584	3.546	3.575
AFFX-LysX-M_at	4.272	4.383	4.305	4.469	4.297	4.260
AFFX-LysX-3_at	3.521	3.580	3.578	3.466	3.504	3.489
AFFX-PheX-5_at	3.832	3.904	3.763	3.717	3.716	3.667
AFFX-PheX-M_at	3.724	3.655	3.567	3.704	3.705	3.595

Recursos - Bases de datos


Gene Expression Omnibus

[GEO Publications](#) | [FAQ](#) | [MIAME](#) | [Email GEO](#)

NCBI » GEO Log

Gene Expression Omnibus: a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles. [More information »](#)

GEO navigation

QUERY

DataSets

Gene profiles

GEO accession

GEO BLAST

GO

GO

GO

BROWSE

DataSets

GEO accessions


Platforms

Samples

Series

Site contents

Public data

Platforms	9,304
Samples	616,735
Series 	24,810
DataSets	2,720

Documentation

[Overview](#) | [FAQ](#) | [Find](#)
[Submission guide](#)
[Linking & citing](#)
[Journal citations](#)
[Construct a Query](#)
[Programmatic access](#)
[DataSet clusters](#)
[GEO announce list](#)
[Data disclaimer](#)

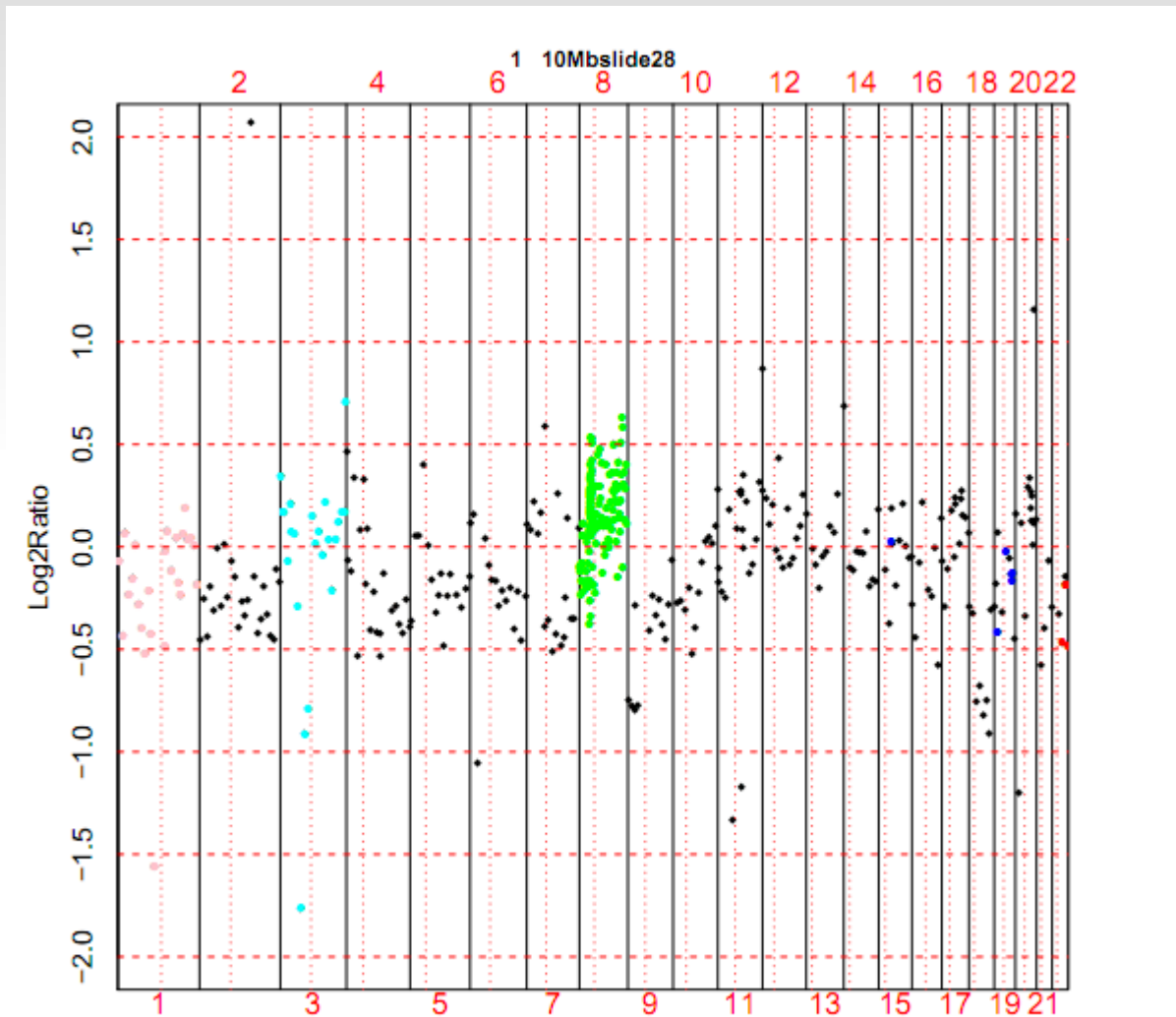
Algunas ideas clave

**Diferentes niveles de información biológica,
distintas situaciones de interés:**

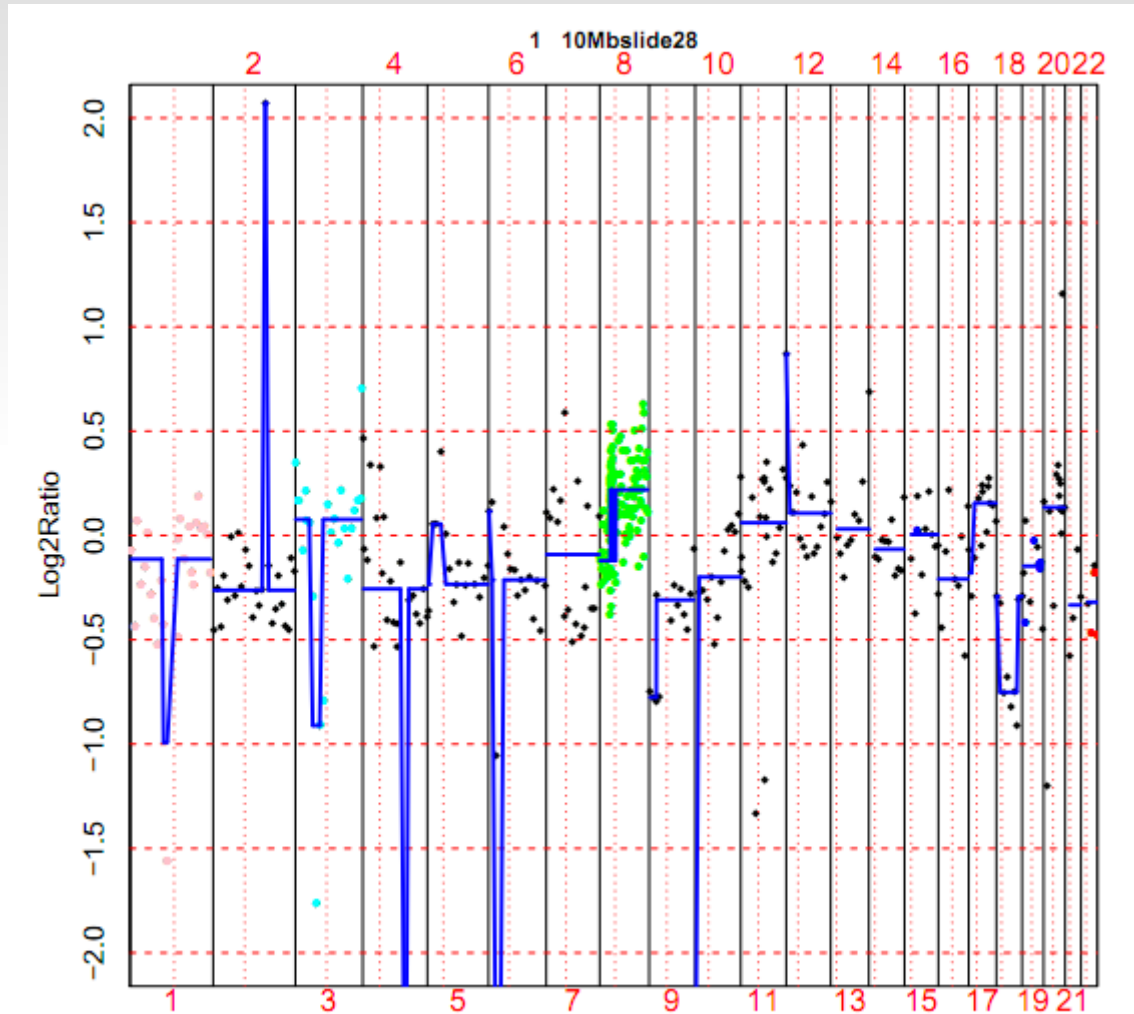
- 1. Variaciones genómicas: CNV, SNP's**
- 2. Expresión génica**
- 3. Interacciones entre proteínas**

extra-slides

Análisis de segmentación



Análisis de segmentación



Análisis de segmentación

